Robust smoothed canonical correlation analysis for functional data

Graciela Boente[1,3] and Nadia L. Kudraszow[2,3]

[1] *Universidad de Buenos Aires,* [2] *Universidad Nacional de La Plata and* [3] *CONICET, Argentina*

Supplementary Material

This supplement contains some comments regarding the Fisher–consistency and interpretation of the robust canonical directions as well as the proofs of Lemma 2 and Proposition 1. It also contains some complementary results regarding the simulation study reported in Section 5. Formulas along this supplement are numbered as (S.1.1), (S.1.2), (S.2.1), (S.2.2) etc according to the Section number. References to sections and equations on the main body of the paper are indicated without the capital "S".

S.1 Regarding C1 and Fisher–consistency

Assume that $(X, Y)^\top \sim \mathcal{E}(\mu, \mathbf{\Gamma}, \varphi)$ where $\mathcal{E}(\boldsymbol{\mu}, \mathbf{\Gamma}, \varphi)$ denotes an elliptical distribution, as defined in Bali and Boente (2009), with parameters

$\mu = (\mu_1, \mu_2)^\top \in \mathcal{H} \times \mathcal{H}$ and $\mathbf{\Gamma}$ is as in (4.1). As a consequence of the definition of elliptical random elements, for any $u, v \in \mathcal{H}$, $(\langle u, X\rangle, \langle v, Y\rangle)^\top \sim \mathcal{E}_2(\boldsymbol{\mu}_{u,v}, \mathbf{\Sigma}_{u,v}, \varphi)$, where $\boldsymbol{\mu}_{u,v} = (\langle u, \mu_1\rangle, \langle v, \mu_2\rangle)^\top$ and the diagonal elements of $\mathbf{\Sigma}_{u,v}$ are $\langle u, \mathbf{\Gamma}_{11} u\rangle$ and $\langle v, \mathbf{\Gamma}_{22} v\rangle$, while the cross-diagonal ones are $\langle u, \mathbf{\Gamma}_{12} v\rangle$ and $\langle v, \mathbf{\Gamma}_{21} u\rangle$. As mentioned in Remark 2, when considering a robust scale functional $\sigma_{\mathrm{R}}(\cdot)$, there exists a constant $c > 0$ such that $\sigma_X^2(u) = c\,\langle u, \mathbf{\Gamma}_{11} u\rangle$ and $\sigma_Y^2(v) = c\,\langle v, \mathbf{\Gamma}_{22} v\rangle$, for any $u, v \in \mathcal{H}$, as required in **C1**. Let us denote ρ_{R} the measure of association $\rho_{\mathrm{R}}(U, V) = \gamma_{\mathrm{R}}(U, V)/\{\sigma_{\mathrm{R}}(U)\,\sigma_{\mathrm{R}}(V)\}$. If ρ_{R} is Fisher-consistent at the family of bivariate elliptical distributions, we have that $\mathcal{L}_{\mathrm{R}}(u, v) = \rho_{\mathrm{R}}^2(P_{(X,Y)}[u, v]) = \mathcal{L}_{\mathbf{\Gamma}}(u, v)$, where

$$\mathcal{L}_{\mathbf{\Gamma}}(u, v) = \frac{\langle u, \mathbf{\Gamma}_{12} v\rangle^2}{\langle u, \mathbf{\Gamma}_{11} u\rangle\,\langle v, \mathbf{\Gamma}_{22} v\rangle}\,, \tag{S.1.1}$$

so that $\gamma_{XY}(u, v) = c\,\langle u, \mathbf{\Gamma}_{12} v\rangle$ which corresponds to the representation in **C1**. On the other side, as mentioned in Remark 2, all association measures in Section 2.2 are Fisher-consistent for bivariate elliptical families, so if the co–association measure γ_{R} is defined through an association measure and a scale functional, the representations given in **C1** and (S.1.1) also hold.

A direct consequence of (S.1.1), is that the estimated canonical directions defined in (3.1) are consistent for the first canonical directions associated to $\mathbf{\Gamma}$, that is, to the maximizers of $\mathcal{L}_{\mathbf{\Gamma}}(u, v)$, which do not depend on ρ_{R}, a fact that was already mentioned in Remark 4.3 from Alvarez,

Boente and Kudraszow (2019) and that states that the robust canonical directions are Fisher-consistent at elliptical processes. In particular, if the process has second moments, taking into account that there exists an $a \in \mathbb{R}$, $a > 0$, such that the covariance operator of $(X, Y)^\top$ equals $a\mathbf{\Gamma}$, we have that $\mathcal{L}_{\text{R}}(u, v) = \mathcal{L}_{\mathbf{\Gamma}}(u, v) = \mathcal{L}(u, v)$, so the robust functionals related to the canonical analysis are the usual ones.

S.2 Interpretation of the robust canonical directions

As mentioned in Section S.1, for elliptical processes with finite second moments and Fisher–consistent association measures, the first robust canonical directions and the robust maximal association equal the usual ones. An important topic is the interpretation that these target quantities have when second moments do not exist. To have an insight on this topic, assume that $\mathcal{L}_{\text{R}}(u, v) = \mathcal{L}_{\mathbf{\Gamma}}(u, v)$, with $\mathcal{L}_{\mathbf{\Gamma}}$ defined in (S.1.1) and $\mathbf{\Gamma}$ a Hilbert-Schmidt operator with finite trace, a property that holds for instance if $(X, Y)^\top \sim \mathcal{E}(\mu, \mathbf{\Gamma}, \varphi)$ and ρ_{R} is a Fisher–consistent association measure. As it will be shown in the sequel, even when second moment do not exist, the maximizers of $\mathcal{L}_{\text{R}}(u, v)$ can be derived using similar arguments to those considered in He, Müller and Wang (2003), providing a framework to

characterize the target directions.

From now on, for $j = 1, 2$, denote as $\{\varphi_\ell^{(j)}\}_{\ell \geq 1}$ the orthonormal basis of eigenfunctions of $\mathbf{\Gamma}_{jj}$ with related eigenvalues $\lambda_1^{(j)} \geq \lambda_2^{(j)} \geq \ldots$, that is, $\mathbf{\Gamma}_{jj} = \sum_{\ell \geq 1} \lambda_\ell^{(j)} \, \varphi_\ell^{(j)} \otimes \varphi_\ell^{(j)}$.

Assume that, for $j = 1, 2$, $\lambda_\ell^{(j)} > 0$ for any $\ell \geq 1$,

$$\sum_{i,j \geq 1} \frac{1}{\lambda_i^{(1)}} r_{ij}^2 < \infty \qquad \text{and} \qquad \sum_{i,j \geq 1} \frac{1}{\lambda_j^{(2)}} r_{ij}^2 < \infty \,, \tag{S.2.1}$$

where

$$r_{\ell s}^2 = \frac{\langle \varphi_\ell^{(1)} \,, \mathbf{\Gamma}_{12} \, \varphi_s^{(2)} \rangle^2}{\lambda_\ell^{(1)} \lambda_s^{(2)}} = \mathcal{L}_{\mathrm{R}}(\varphi_\ell^{(1)}, \varphi_s^{(2)}) \,.$$

Note that (S.2.1) is analogous to condition 4.5 in He, Müller and Wang (2003) and clearly entails that $\sum_{i,j \geq 1} r_{ij}^2 < \infty$, which corresponds to equation (15) from the same paper. For $j = 1, 2$, define the Hilbert-Schmidt operators $\mathbf{R}_{jj} = \sum_{\ell \geq 1} \sqrt{\lambda_\ell^{(j)}} \, \varphi_\ell^{(j)} \otimes \varphi_\ell^{(j)}$ with range

$$\mathcal{R}_{jj} = \left\{ y \in \mathcal{H} : \quad \sum_{\ell \geq 1} \frac{1}{\lambda_\ell^{(j)}} \, \langle y, \phi_\ell^{(1)} \rangle < \infty \right\} .$$

Therefore, the inverse of $\mathbf{R}_{jj}$ is well defined over $\mathcal{R}_{jj}$ as

$$\mathbf{R}_{jj}^{-1} = \sum_{\ell \geq 1} \frac{1}{\sqrt{\lambda_\ell^{(j)}}} \, \varphi_\ell^{(j)} \otimes \varphi_\ell^{(j)} \,.$$

Furthermore, denote as $\mathbf{\Upsilon}_{12} = \mathbf{R}_{11}^{-1} \mathbf{\Gamma}_{12} \mathbf{R}_{22}^{-1}$ and $\mathbf{\Upsilon}_{21} = \mathbf{R}_{22}^{-1} \mathbf{\Gamma}_{21} \mathbf{R}_{11}^{-1}$.

Using similar arguments to those considered in the proof of Proposition 4.2 in Horváth and Kokoszka (2012), we get that under (S.2.1),

$\boldsymbol{\Upsilon}_{12} : \mathcal{R}_{22} \to \mathcal{R}_{11}$, $\boldsymbol{\Upsilon}_{21} : \mathcal{R}_{11} \to \mathcal{R}_{22}$ and $\boldsymbol{\Upsilon}_{21} = \boldsymbol{\Upsilon}_{12}^*$. Thus, the operators $\boldsymbol{\Upsilon}_1 = \boldsymbol{\Upsilon}_{12}\ \boldsymbol{\Upsilon}_{12}^* = \mathbf{R}_{11}^{-1}\boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_{22}^{-1}\boldsymbol{\Gamma}_{21}\mathbf{R}_{11}^{-1}$ and $\boldsymbol{\Upsilon}_2 = \mathbf{R}_{22}^{-1}\boldsymbol{\Gamma}_{21}\boldsymbol{\Gamma}_{11}^{-1}\ \boldsymbol{\Gamma}_{12}\mathbf{R}_{22}^{-1}$ are self-adjoint, positive definite, Hilbert-Schmidt operators with the same eigenvalues ν_k^2. Let $\{\omega_\ell^{(j)}\}_{\ell\geq 1}$ the orthonormal basis of eigenfunctions of $\boldsymbol{\Upsilon}_j$. Then, we have that $\omega_\ell^{(1)} = (1/\nu_\ell)\ \boldsymbol{\Upsilon}_{12}(\omega_\ell^{(2)})$ and $\omega_\ell^{(2)} = (1/\nu_\ell)\ \boldsymbol{\Upsilon}_{21}(\omega_\ell^{(1)})$.

Define $\phi_k = \mathbf{R}_{11}^{-1}\omega_\ell^{(1)}$ and $\psi_k = \mathbf{R}_{22}^{-1}\omega_\ell^{(2)}$, hence we easily get that $\langle \phi_k, \boldsymbol{\Gamma}_{12}\,\psi_k\rangle = \nu_k$, $\langle \phi_k, \boldsymbol{\Gamma}_{11}\phi_k\rangle = 1$, $\langle \psi_k, \boldsymbol{\Gamma}_{22}\psi_k\rangle = 1$. Moreover, for any $u, v \in \mathcal{H}$, such that $\langle u, \boldsymbol{\Gamma}_{11}u\rangle = \langle v, \boldsymbol{\Gamma}_{22}v\rangle = 1$ and $\langle u, \boldsymbol{\Gamma}_{11}\phi_j\rangle = 0$, $\langle v, \boldsymbol{\Gamma}_{22}\psi_j\rangle = 0$ for any $j < k$, we obtain that $\langle u, \boldsymbol{\Gamma}_{12}v\rangle \leq \nu_k$, so assumption **C2** holds taking $\rho_1 = \nu_1$, $\rho_2 = \nu_2$ if $\nu_2 < \nu_1$. Note that the obtained expressions provide an interpretation of ρ_1, ϕ_1 and ψ_1 analogous to those given in classical FCCA, but in terms of the scatter operator $\boldsymbol{\Gamma}$.

S.3 Proofs of Lemma 2 and Proposition 1

Proof of Lemma 2. To show that $\mathcal{L}_{\mathrm{R}}(u, v) \geq \mathcal{L}_{\tau,\mathrm{R}}(u, v)$ note that, given any u and v in $\mathcal{H}_{\mathrm{s}}$

$$\frac{\mathcal{L}_{\tau,\mathrm{R}}(u, v)}{\mathcal{L}_{\mathrm{R}}(u, v)} = \frac{\sigma_X^2(u)}{\{\sigma_X^2(u) + \tau\Psi(u)\}}\frac{\sigma_Y^2(v)}{\{\sigma_Y^2(v) + \tau\Psi(v)\}} \leq 1 \tag{S.3.1}$$

where the inequality follows from the fact that σ_{R}^2, τ and $\Psi(\cdot)$ are non negative. Thus, $\lambda_\tau = \mathcal{L}_{\tau,\mathrm{R}}(\phi_{\tau,1}, \psi_{\tau,1}) \leq \mathcal{L}_{\mathrm{R}}(\phi_{\tau,1}, \psi_{\tau,1}) \leq \lambda_0$, proving the

second part. From the expression for the ratio $\mathcal{L}_{\tau,\mathrm{R}}(u,v)/\mathcal{L}_{\mathrm{R}}(u,v)$ given in (S.3.1) we conclude that, for any fixed u and v in $\mathcal{H}_{\mathrm{S}}$, $\mathcal{L}_{\tau,\mathrm{R}}(u,v) \to \mathcal{L}_{\mathrm{R}}(u,v)$ as $\tau \to 0$. Hence, using that $\lambda_0 \geq \lambda_\tau \geq \mathcal{L}_{\tau,\mathrm{R}}(\phi_1,\psi_1)$ and the fact that $\mathcal{L}_{\tau,\mathrm{R}}(\phi_1,\psi_1) \to \mathcal{L}_{\mathrm{R}}(\phi_1,\psi_1) = \lambda_0$, we conclude the proof. □

Proof of Proposition 1. From the equivariance of the scale functional and **C4**, it follows that

$$\sup_{u \in \mathcal{H}_{\mathrm{S}}^0} \left| \frac{s_{n,X}^2(u) + \tau\Psi(u)}{\sigma_X^2(u) + \tau\Psi(u)} - 1 \right| = C_{n,X} \xrightarrow{a.s.} 0\,.$$

Similarly, we have that

$$\sup_{v \in \mathcal{H}_{\mathrm{S}}^0} \left| \frac{s_{n,Y}^2(v) + \tau\Psi(v)}{\sigma_Y^2(v) + \tau\Psi(v)} - 1 \right| = C_{n,Y} \xrightarrow{a.s.} 0\,.$$

Note that, if **C4**(a) holds,

$$\begin{aligned} \sup_{u,v \in \mathcal{H}_{\mathrm{S}}^0} \frac{|g_n(u,v) - \gamma_{XY}(u,v)|}{\left\{\sigma_X^2(u)\rangle + \tau\Psi(u)\right\}^{1/2} \left\{\sigma_Y^2(v) + \tau\Psi(v)\right\}^{1/2}} &= \sup_{u,v \in \mathcal{H}_{\mathrm{S}}^0} \frac{|g_n(u,v) - \gamma_{XY}(u,v)|}{\|u\|_{1,\tau}\|v\|_{2,\tau}} \\ &= \sup_{\|u\|_{1,\tau}=\|v\|_{2,\tau}=1} |g_n(u,v) - \gamma_{XY}(u,v)| \\ &= C_{n,XY} \xrightarrow{a.s.} 0\,. \end{aligned}$$

Let us first prove our assertion when **C4**(a) holds. Note that

$$
\begin{aligned}
&|\widehat{\mathcal{L}}_{\tau_n,\mathrm{R}}(u,v) - \mathcal{L}_{\tau_n,\mathrm{R}}(u,v)| = \\
&= \left| \frac{g_n^2(u,v)}{\{s_{n,X}^2(u)+\tau\Psi(u)\}\{s_{n,Y}^2(v)+\tau\Psi(v)\}} - \frac{\gamma_{XY}^2(u,v)}{\{\sigma_X^2(u)+\tau\Psi(u)\}\{\sigma_Y^2(v)+\tau\Psi(v)\}} \right| \\
&\leq \frac{g_n^2(u,v)}{\{s_{n,X}^2(u)+\tau\Psi(u)\}\{s_{n,Y}^2(v)+\tau\Psi(v)\}} \left| \frac{\{s_{n,X}^2(u)+\tau\Psi(u)\}\{s_{n,Y}^2(v)+\tau\Psi(v)\}}{\{\sigma_X^2(u)+\tau\Psi(u)\}\{\sigma_Y^2(v)+\tau\Psi(v)\}} - 1 \right| \\
&+ \frac{|g_n^2(u,v) - \gamma_{XY}^2(u,v)|}{\{\sigma_X^2(u)+\tau\Psi(u)\}\{\sigma_Y^2(v)+\tau\Psi(v)\}} = A_{1,n}(u,v) + A_{2,n}(u,v)\,.
\end{aligned}
$$

Let us begin by bounding $A_{2,n}(u,v)$. Taking into account that for any $u, v \in \mathcal{H}_{\mathrm{S}}^0$, we have that

$$
\frac{|\gamma_{XY}^2(u,v)|}{\{\sigma_X^2(u)+\tau\Psi(u)\}\{\sigma_Y^2(v)+\tau\Psi(v)\}} \leq \mathcal{L}_{\mathrm{R}}(u,v) \leq \lambda_0\,,
$$

and using **C4**(a), we obtain that

$$
\begin{aligned}
&\sup_{u,v\in\mathcal{H}_{\mathrm{S}}^0} A_{2,n}(u,v) \leq \\
&\qquad \leq \sup_{u,v\in\mathcal{H}_{\mathrm{S}}^0} \frac{\{g_n(u,v)-\gamma_{XY}(u,v)\}^2 + 2|\gamma_{XY}(u,v)|\ |g_n(u,v)-\gamma_{XY}(u,v)|}{\{\sigma_X^2(u)+\tau\Psi(u)\}\{\sigma_Y^2(v)+\tau\Psi(v)\}} \\
&\qquad \leq C_{n,XY}^2 + \lambda_0^{1/2}\, C_{n,XY} \xrightarrow{a.s.} 0\,.
\end{aligned}
$$

On the other hand, to bound $A_{1,n}(u,v)$, we use that $ab-1 = (a-1)(b-1)+(a-1)+(b-1)$. Choosing $a = \{s_{n,X}^2(u)+\tau\Psi(u)\}/\{\sigma_X^2(u)+\tau\Psi(u)\}$

and $b = \{s^2_{n,Y}(v) + \tau\Psi(v)\}/\{\sigma^2_Y(v) + \tau\Psi(v)\}$, we obtain that

$$\begin{aligned}\sup_{u,v\in\mathcal{H}^0_{\mathrm{S}}} & \left|\frac{\{s^2_{n,X}(u)+\tau\Psi(u)\}\{s^2_{n,Y}(v)+\tau\Psi(v)\}}{\{\sigma^2_X(u)+\tau\Psi(u)\}\{\sigma^2_Y(v)+\tau\Psi(v)\}} - 1\right| \leq \\ & \leq \sup_{u,v\in\mathcal{H}^0_{\mathrm{S}}}\left|\frac{s^2_{n,X}(u)+\tau\Psi(u)}{\sigma^2_X(u)+\tau\Psi(u)} - 1\right| \sup_{u,v\in\mathcal{H}^0_{\mathrm{S}}}\left|\frac{s^2_{n,Y}(v)+\tau\Psi(v)}{\sigma^2_Y(v)+\tau\Psi(v)} - 1\right| \\ & + \sup_{u,v\in\mathcal{H}^0_{\mathrm{S}}}\left|\frac{s^2_{n,X}(u)+\tau\Psi(u)}{\sigma^2_X(u)+\tau\Psi(u)} - 1\right| + \sup_{u,v\in\mathcal{H}^0_{\mathrm{S}}}\left|\frac{s^2_{n,Y}(v)+\tau\Psi(v)}{\sigma^2_Y(v)+\tau\Psi(v)} - 1\right| \\ & \leq C_{n,X}\ C_{n,Y} + C_{n,X} + C_{n,Y} \xrightarrow{a.s.} 0\ .\end{aligned}$$

Using that, for any $u, v\ \in \mathcal{H}^0_{\mathrm{S}}$, $\widehat{\mathcal{L}}_{\tau,\mathrm{R}}(u,v) \leq A$, we get that $\sup_{u,v\in\mathcal{H}^0_{\mathrm{S}}} A_{1,n}(u,v) \xrightarrow{a.s.}$ 0, which concludes the proof when **C4**(a) holds.

Assume now that **C4**(b) holds. Then, $\gamma^2_n(u,v) = r^2_n(u,v)s^2_{n,X}(u)s^2_{n,Y}(v)$, $\gamma^2_{XY}(u,v) = \rho^2_{XY}(u,v)\sigma^2_X(u)\sigma^2_Y(v)$. Thus, for any $u, v \in \mathcal{H}^0_{\mathrm{S}}$,

$$|\widehat{\mathcal{L}}_{\tau_n,\mathrm{R}}(u,v) - \mathcal{L}_{\tau_n,\mathrm{R}}(u,v)| \leq B_{1,n}(u,v) + B_{2,n}(u,v)\,,$$

where

$$B_{1,n}(u,v) = \frac{|r^2_n(u,v) - \rho^2_{XY}(u,v)|\, s^2_{n,X}(u)s^2_{n,Y}(v)}{\{s^2_{n,X}(u)+\tau\Psi(u)\}\{s^2_{n,Y}(v)+\tau\Psi(v)\}} \qquad \text{and}$$

$$B_{2,n}(u,v) = \left|\frac{s^2_{n,X}(u)s^2_{n,Y}(v)}{\{s^2_{n,X}(u)+\tau\Psi(u)\}\{s^2_{n,Y}(v)+\tau\Psi(v)\}} - \frac{\sigma^2_X(u)\sigma^2_Y(v)}{\{\sigma^2_X(u)+\tau\Psi(u)\}\{\sigma^2_Y(v)+\tau\Psi(v)\}}\right| .$$

Using that $r^2_n(u,v) \leq 1$ and $\rho^2_{XY}(u,v) \leq 1$, from **C4**(b) we immediately obtain that $B_{1,n} \leq 2\ \sup_{\|u\|=\|v\|=1} |r_n(u,v) - \rho_{XY}(u,v)| = 2\ \theta_n \xrightarrow{a.s.} 0$. It

only remains to show that $\sup_{u,v\in\mathcal{H}_{\mathrm{S}}^0} B_{2,n}(u,v) \xrightarrow{a.s.} 0$. Note that

$$\begin{aligned}\sup_{u,v\in\mathcal{H}_{\mathrm{S}}^0} B_{2,n}(u,v) \leq & \sup_{u,v\in\mathcal{H}_{\mathrm{S}}^0} \frac{s_{n,X}^2(u)s_{n,Y}^2(v)}{s_{n,X}^2(u)+\tau\Psi(u)} \left| \frac{1}{s_{n,Y}^2(v)+\tau\Psi(v)} - \frac{1}{\sigma_Y^2(v)+\tau\Psi(v)} \right| \\ & + \sup_{u,v\in\mathcal{H}_{\mathrm{S}}^0} \left| \frac{s_{n,X}^2(u)s_{n,Y}^2(v)}{\{s_{n,X}^2(u)+\tau\Psi(u)\}\{\sigma_Y^2(v)+\tau\Psi(v)\}} - \right. \\ & \left. - \frac{\sigma_X^2(u)\sigma_Y^2(v)}{\{\sigma_X^2(u)+\tau\Psi(u)\}\{\sigma_Y^2(v)+\tau\Psi(v)\}} \right| \end{aligned}$$

$$\begin{aligned} & \leq C_{n,Y} \sup_{u,v\in\mathcal{H}_{\mathrm{S}}^0} \frac{s_{n,X}^2(u)}{s_{n,X}^2(u)+\tau\Psi(u)} \frac{s_{n,Y}^2(v)}{s_{n,Y}^2(v)+\tau\Psi(v)} \\ & + \sup_{u,v\in\mathcal{H}_{\mathrm{S}}^0} \left| \frac{s_{n,X}^2(u)s_{n,Y}^2(v)}{\{s_{n,X}^2(u)+\tau\Psi(u)\}\{\sigma_Y^2(v)+\tau\Psi(v)\}} - \right. \\ & \qquad \left. - \frac{\sigma_X^2(u)\sigma_Y^2(v)}{\{\sigma_X^2(u)+\tau\Psi(u)\}\{\sigma_Y^2(v)+\tau\Psi(v)\}} \right| \\ & \leq C_{n,Y} + \sup_{u,v\in\mathcal{H}_{\mathrm{S}}^0} \left| \frac{s_{n,X}^2(u)s_{n,Y}^2(v)}{\{s_{n,X}^2(u)+\tau\Psi(u)\}\{\sigma_Y^2(v)+\tau\Psi(v)\}} - \right. \\ & \qquad \left. - \frac{\sigma_X^2(u)\sigma_Y^2(v)}{\{\sigma_X^2(u)+\tau\Psi(u)\}\{\sigma_Y^2(v)+\tau\Psi(v)\}} \right| . \end{aligned}$$

Denote $D_{1,n}$ the second term on the right hand side of the above equation.

Arguing similarly we get that

$$
\begin{aligned}
D_{1,n} &\leq C_{n,X} \sup_{u,v \in \mathcal{H}_S^0} \frac{s^2_{n,Y}(v)}{\sigma^2_Y(v) + \tau\Psi(v)} + \sup_{u,v \in \mathcal{H}_S^0} \frac{\left|s^2_{n,X}(u)s^2_{n,Y}(v) - \sigma^2_X(u)\sigma^2_Y(v)\right|}{\{\sigma^2_X(u) + \tau\Psi(u)\}\{\sigma^2_Y(v) + \tau\Psi(v)\}} \\
&\leq C_{n,X} + C_{n,X} \sup_{u,v \in \mathcal{H}_S^0} \frac{|s^2_{n,Y}(v) - \sigma^2_Y(v)|}{\sigma^2_Y(v) + \tau\Psi(v)} + \\
&+ \sup_{u,v \in \mathcal{H}_S^0} \frac{\left|s^2_{n,X}(u)s^2_{n,Y}(v) - \sigma^2_X(u)\sigma^2_Y(v)\right|}{\{\sigma^2_X(u) + \tau\Psi(u)\}\{\sigma^2_Y(v) + \tau\Psi(v)\}} \\
&\leq C_{n,X} + C_{n,X}\, C_{n,Y} + \sup_{u,v \in \mathcal{H}_S^0} \frac{\left|s^2_{n,X}(u)s^2_{n,Y}(v) - \sigma^2_X(u)\sigma^2_Y(v)\right|}{\{\sigma^2_X(u) + \tau\Psi(u)\}\{\sigma^2_Y(v) + \tau\Psi(v)\}} \\
&= C_{n,X} + C_{n,X}\, C_{n,Y} + D_{2,n}\,.
\end{aligned}
$$

Finally $D_{2,n}$ can easily be bounded by $C_{n,X}\, C_{n,Y} + C_{n,X} + C_{n,Y}$ so

$$
\sup_{u,v \in \mathcal{H}_S^0} B_{2,n}(u,v) \leq C_{n,Y} + 2\,(C_{n,X} + C_{n,X}\, C_{n,Y}) \xrightarrow{a.s.} 0\,.
$$

concluding the proof. □

S.4 Monte Carlo study

This Section contains some additional results to those reported in Section 5.1. As mentioned therein, we consider the robust association measure induced from a bivariate robust $M-$scatter functional, described in Section 2.2, computed using Huber's score function with tuning constant $k_1 = (\chi^2_{2,0.9})^{1/2}$. From now on, the classical and robust estimators will be labeled as CL and ROB, respectively.

S.4.1 Simulation settings

Our simulation model is similar to the one considered in He, Müller and Wang (2004) and Alvarez, Boente and Kudraszow (2019). For each replication, we generate independent samples $\{(X_i, Y_i)^\top\}_{i=1}^n \subset \mathcal{H} \times \mathcal{H}$ of size $n = 100$ with $\mathcal{H} = L^2[0, 50]$. The processes are observed over an equispaced grid of 50 points t_j, $j = 1, \ldots, 50$.

For clean data sets, (X_i, Y_i) are generated with the same distribution as the Gaussian random element $(X, Y)^\top \in \mathcal{H} \times \mathcal{H}$, given by $X(t) = \sum_{j=1}^m \eta_j f_j(t)$ and $Y(t) = \sum_{j=1}^m \zeta_j f_j(t)$, where $\{f_j\}_{j\geq 1}$ is the Fourier basis of $L^2[0, 50]$ and $m = 21$. The scores $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_m)^\top$ and $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_m)^\top$ are $m-$dimensional normally distributed random vectors, $(\boldsymbol{\eta}^\top, \boldsymbol{\zeta}^\top)^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with the diagonal blocks of the matrix $\boldsymbol{\Sigma}$ equal to $\boldsymbol{\Sigma}_{22} = \boldsymbol{\Sigma}_{11} = 10\,\mathrm{diag}(1, 1, 1, 0.75, \ldots, 0.75^{m-3})$, while the off–diagonal one equals $\boldsymbol{\Sigma}_{12} = \mathrm{diag}(7, 3, 1, 0, \ldots, 0)$. This setting will be labelled C_0.

As described in the main document, two contamination schemes are studied denoted C_1 and C_2, the first one being a strong contamination in the direction of the second canonical direction of $(X, Y)^\top$, while in the latter, we contaminate in the direction of a linear combination of the third canonical weight and the fourth element of the basis which, for clean sam-

ples, corresponds to a canonical direction with null canonical correlation. The contaminated trajectories are denoted $(X_i^{(c)}, Y_i^{(c)})^\top$. For the sake of completeness, we recall how the contaminated curves are obtained.

- Under C_1, $(X_i^{(c)}, Y_i^{(c)})^\top$ are i.i.d. with the same distribution as $(1 - B)(X, Y)^\top + BW(f_2, f_2)^\top$, where $B \sim \mathcal{B}(1, 0.1)$, $W \sim \mathcal{N}(25, 1)$ and W, B, $(X, Y)^\top$ are all independent of each other.

- When considering the scheme C_2, $(X_i^{(c)}, Y_i^{(c)})^\top$ are i.i.d. and such that

$$\begin{aligned} X_i^{(c)} &\sim (1 - B)\, X + B \left(\eta_1\, f_1 + W \frac{f_3 + f_4}{\sqrt{2}} + 0.1\, \eta_3\, f_3 + 0.1\, \eta_4\, f_4 \right) \\ Y_i^{(c)} &\sim (1 - B)\, Y + B \left(\zeta_1\, f_1 + W \frac{f_3 + f_4}{\sqrt{2}} + 0.1\, \zeta_3\, f_3 + 0.1\, \zeta_4\, f_4 \right) \end{aligned}$$

  with $W \sim \mathcal{N}(25, 0.01)$ independent of $B \sim \mathcal{B}(1, 0.1)$, X, Y and $(\boldsymbol{\xi}^\top, \boldsymbol{\zeta}^\top)^\top \sim \mathcal{N}(0, \boldsymbol{\Sigma})$.

In all cases, we performed $NR = 1000$ replications. Figure S.1 illustrates one of the samples obtained when no contamination is introduced and under each contamination setting. It shows how the introduced contaminations modify the pattern of the clean data when $\epsilon = 0.10$. The red lines correspond to the introduced outliers.

Two basis $\{\xi_i\}_{i \geq 1}$ are considered: the cubic $B-$spline basis and the Fourier basis and the dimension of the linear space equals $d = 5, 9, 13$ for

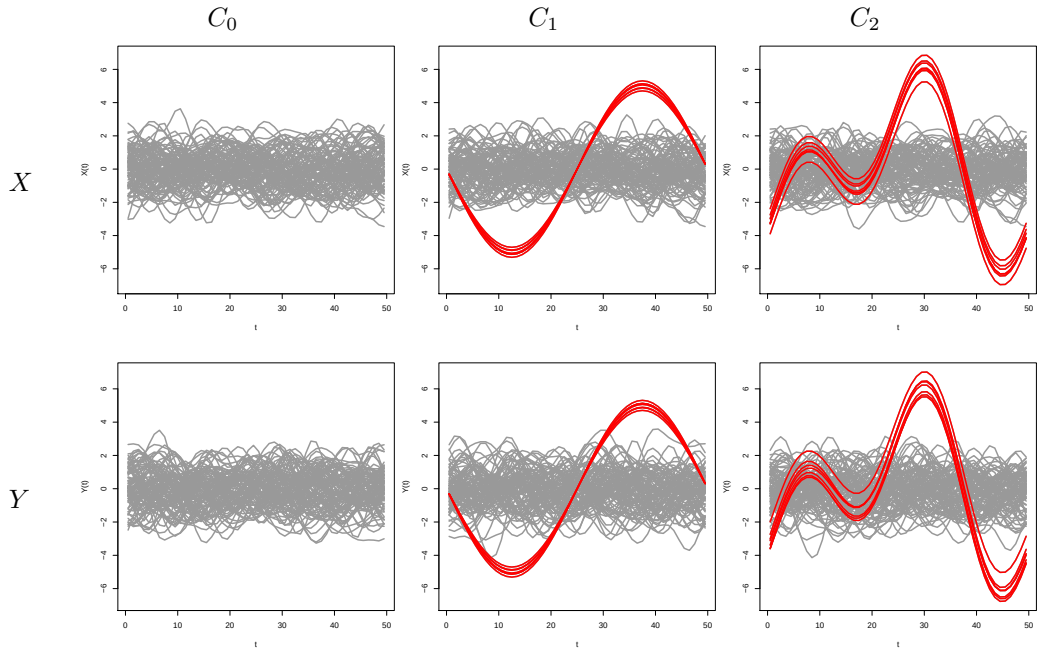

Figure S.1: One typical sample from each contamination setting, with atypical trajectories highlighted in red.

the Fourier basis and $d = 5, 9, 13, 20$ for the $B-$spline one.

The selected penalization operator is the L^2 norm of the second derivative, that is, $\Psi(u) = \int_0^{50} u''(t)^2 dt$. As in Leurgans, Moyeed and Silverman (1993), the discretization of the roughness penalty is computed over the same design points $\{t_j\}_{1 \leq j \leq 50}$, and the second derivative of u at t_i is approximated by $\{u(t_{i+1}) - 2u(t_i) + u(t_{i-1})\}/(t_{i+1} - t_i)^2$, where $u(t_0) = u(t_{50})$ and $u(t_{51}) = u(t_1)$. The values $\lceil \xi_i, \xi_j \rceil = \int_0^{50} \xi_i''(t)\, \xi_j''(t) dt$ for $i, j = 1, \ldots, d$ are evaluated using the approximate discretization of the second derivative

in each element of the basis and then approximating the integral by sums. The values of the smoothing parameter are $\tau = 20, 30, 40$.

The smoothed estimators $(\widetilde{\phi}_{\boldsymbol{\kappa},1}, \widetilde{\psi}_{\boldsymbol{\kappa},1})$ given by (3.2) are computed using the algorithm described in Section S.4.2, while the proposal $(\widehat{\phi}_{\tau,1}, \widehat{\psi}_{\tau,1})$ defined in (3.1) and obtained without projecting the data on a basis, is obtained with the same algorithm but replacing the products and norms by approximations of the inner products and norms in $\mathcal{H}$.

S.4.2 The algorithm

Given $\{\xi_i\}_{i\geq 1}$ an orthonormal basis for $\mathcal{H}$, the estimators proposed in (3.2) are obtained searching directions $u = \sum_{i=1}^{d} \alpha_i \xi_i \in \mathcal{H}_d^0$ and $v = \sum_{i=1}^{d} \beta_i \xi_i \in \mathcal{H}_d^0$ that lead to the maximum value of $\widehat{\mathcal{L}}_{\boldsymbol{\tau},\mathrm{R}}(u, v)$. In order to achieve identifiability up to a sign the norms are set to 1, i.e. $\sum_{i=1}^{d} \alpha_i^2 = \sum_{i=1}^{d} \beta_i^2 = 1$. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^\top$ be the coefficients' vectors of u and v in the considered basis and denote $\mathbf{x} = (\langle X, \xi_1 \rangle, \ldots, \langle X, \xi_d \rangle)^\top$ and $\mathbf{y} = (\langle Y, \xi_1 \rangle, \ldots, \langle Y, \xi_d \rangle)^\top$. Noting that $\langle u, X \rangle = \boldsymbol{\alpha}^\top \mathbf{x}$ and $\langle v, Y \rangle = \boldsymbol{\beta}^\top \mathbf{y}$, it is easily seen that the robust SCCA, in the basis expansion domain, can be obtained using any multivariate algorithm allowing to find the vectors $\widehat{\boldsymbol{\alpha}}_1 = (\widehat{\alpha}_{11}, \ldots, \widehat{\alpha}_{1d})^\top$ and $\widehat{\boldsymbol{\beta}}_1 = (\widehat{\beta}_{11}, \ldots, \widehat{\beta}_{1d})^\top$ with norm 1 which maximize $\widehat{\mathcal{L}}^d_{\boldsymbol{\tau},\mathrm{R}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$, where $\widehat{\mathcal{L}}^d_{\boldsymbol{\tau},\mathrm{R}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is defined by replacing the sam-

ple covariance and variances in (2.3) by sample versions of the robust co–association and scale functionals. The estimations of the coefficients, $\widehat{\boldsymbol{\alpha}}_1$ and $\widehat{\boldsymbol{\beta}}_1$, were computed adapting the algorithm proposed by Branco *et al.* (2005) for robust multivariate canonical correlation analysis based on projection–pursuit. The aim is to maximize the projection index $\widehat{\mathcal{L}}^d_{\boldsymbol{\tau},\mathrm{R}}(\boldsymbol{\alpha},\boldsymbol{\beta})$ under the restrictions $\boldsymbol{\alpha}^\top\boldsymbol{\alpha}=1$ and $\boldsymbol{\beta}^\top\boldsymbol{\beta}=1$. To get rid of the restrictions on the norms, it is convenient to write the arguments $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of $\widehat{\mathcal{L}}^d_{\boldsymbol{\tau},\mathrm{R}}$ in polar coordinates. Let $(\theta_1,\ldots,\theta_{d-1})^\top$ be the polar coordinates of a unit norm vector, then $\widehat{\mathcal{L}}^d_{\boldsymbol{\tau},\mathrm{R}}$ is maximized over two sets of $(d-1)$ angles corresponding to each of the arguments $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ with norms equal to one. A standard maximization routine is used and the maximization is performed on a hyperrectangle. Once the optimal angles $(\widehat{\theta}_1,\ldots,\widehat{\theta}_{d-1})^\top$ are obtained, $\widehat{\boldsymbol{\alpha}}_1$ and $\widehat{\boldsymbol{\beta}}_1$ are computed going back to the original coordinates using a recursive relation (see Branco *et al.* (2005) for further details). Finally, the canonical direction estimators in the basis expansion domain can be reconstructed as $\widetilde{\phi}_{\boldsymbol{\kappa},1}=\sum_{j=1}^d\widehat{\alpha}_{1j}\xi_j$ and $\widetilde{\psi}_{\boldsymbol{\kappa},1}=\sum_{j=1}^d\widehat{\beta}_{1j}\xi_j$, with $\boldsymbol{\kappa}=(\tau,d)$.

S.4.3 Results

For each situation, to evaluate the performance of a given estimator $(\widehat{\phi}_1,\widehat{\psi}_1$ of the first canonical weights (ϕ_1,ψ_1), we compute

- the average over replications of $\|\widehat{\phi}_1-\phi_1\|^2+\|\widehat{\psi}_1-\psi_1\|^2$, denoted MISE.

- the average over replications of the absolute Pearson correlation between the canonical variates computed only for the non–atypical data, that is, the average of

$$\widehat{\rho}_X = \widehat{\rho}_{\text{CL},XX,\text{CLEAN}}(\phi_1,\widehat{\phi}_1) = \frac{\left|\sum_{i=1}^n (1-B_i)U_i\widehat{U}_i\right|}{\sqrt{\sum_{i=1}^n (1-B_i)\widehat{U}_i^2 \sum_{i=1}^n (1-B_i)U_i^2}},$$

where $U_i = \langle X_i, \phi_1\rangle$ and $\widehat{U}_i = \langle X_i, \widehat{\phi}_1\rangle$. This measure provides a way to quantify how the proposals fit the good observations. A similar measure was computed for $\widehat{\psi}_1$.

Tables S.1 and S.2 report the obtained results when considering the Fourier and $B-$spline bases, respectively, while Table S.3 summarizes the results when considering the estimators $(\widehat{\phi}_{\tau,1}, \widehat{\psi}_{\tau,1})$ defined in (3.1). Besides, taking into account that the quantities $\widehat{\rho}_X$ and $\widehat{\rho}_Y$ are non-negative and expected to have a skewed distribution, we present skewed-adjusted boxplots (see Hubert and Vandervieren (2008)) to display the obtained results in Figures S.2 to S.4. The red and blue boxes correspond to the classical and robust procedures, respectively, when considering the penalized approach based on basis reduction.

Regarding the estimators defined in the basis expansion domain, similarly to the comments given in the main document regarding the MISE,

when no outliers are present all procedures are comparable, and the average values of $\widehat{\rho}_X$ and $\widehat{\rho}_Y$ are close to 1. This fact is more evident in Figures S.3 and S.4 which reveals that under C_0 all the procedures lead to similar results. On the other hand, for contaminated samples, the classical estimators are strongly affected by the presence of outliers. Indeed, in this case Tables S.1 and S.2 show that the average values of $\widehat{\rho}_X$ and $\widehat{\rho}_Y$ move away from 1 approaching 0 (see also Figures S.3 and S.4). On the other hand, the robust estimators give more resistant results independently of the basis choice and the values of the parameters.

The procedure based only on penalization defined through (3.1) leads to values of the average MISE which, in all cases, are larger than those obtained with the basis expansion approach, a fact that was already noticed for the classical estimators and clean samples in He, Müller and Wang (2004). For this family of procedures, the classical method is still highly affected by the considered contaminations models, while the robust one is more stable (see Table S.3). Moreover, the robust estimators perform under C_1 and C_2 similarly to the worst scenario when considering the robust smoothed estimators on the basis expansion domain. Hence, for the considered situations, using robust association and dispersion measure and projecting on a basis before penalizing seems to provide the best choice.

Acknowledgements The authors wish to thank the Associate Editor and two anonymous referees for valuable comments which led to an improved version of the original paper. This research was partially supported by Grants PICT 2018-00740 from ANPCYT, 20020170100022BA from the Universidad de Buenos Aires at Buenos Aires, Argentina (Graciela Boente), the Spanish Project MTM2016-76969P from the Ministry of Economy and Competitiveness, Spain (MINECO/ AEI/FEDER, UE) (Graciela Boente) and PPID X030 and PID I231 from Universidad Nacional de La Plata, Argentina (Nadia Kudraszow).

References

Alvarez, A., Boente, G. and Kudraszow, N. (2019). Robust sieve estimators for functional canonical correlation analysis. *Journal of Multivariate Analysis*, **170**, 46-62.

Bali L. and Boente G. (2009). Principal points and elliptical distributions from the multivariate setting to the functional case. *Statistics and Probability Letters*, **79**, 1858-1865.

Branco, J. A., Croux, C., Filzmoser, P. and Oliveira, M. R. (2005). Robust canonical correlations: A comparative study. *Computational Statistics*, **20**, 203-229.

He, G., Müller, H. G. and Wang, J. L. (2003). Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis*, **85**, 54-77.

He, G., Müller, H. G. and Wang, J. L. (2004). Methods of canonical analysis for functional data. *Journal of Statistical Planning and Inference*, **122**, 141-159.

Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*.

Springer, New York.

Hubert, M. and Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis*, **52**, 5186-5201.

Leurgans, S. E., Moyeed, R. A. and Silverman, B. W. (1993). Canonical correlation analysis when the data are curves, *Journal of the Royal Society, Series B*, **55**, 725-740.

Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Instituto de Cálculo-CONICET, Ciudad Universitaria, Pabellón 2, 1428, Buenos Aires, Argentina

E-mail: (gboente@dm.uba.ar)

Departamento de Matemática, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, CMaLP-CONICET, Casilla de Correo 172, (1900) La Plata, Argentina

E-mail: (nkudraszow@mate.unlp.edu.ar)

Penalty parameter		$\tau = 20$			$\tau = 30$			$\tau = 40$		
d	Estimator	MISE	$\widehat{\rho}_X$	$\widehat{\rho}_Y$	MISE	$\widehat{\rho}_X$	$\widehat{\rho}_Y$	MISE	$\widehat{\rho}_X$	$\widehat{\rho}_Y$
					C_0					
5	CL	0.13	0.97	0.97	0.13	0.97	0.97	0.13	0.97	0.97
5	ROB	0.18	0.96	0.96	0.19	0.96	0.96	0.18	0.96	0.96
	CL	0.31	0.96	0.96	0.27	0.96	0.96	0.25	0.96	0.96
9	ROB	0.42	0.94	0.94	0.37	0.95	0.95	0.34	0.95	0.95
13	CL	0.33	0.96	0.96	0.29	0.96	0.96	0.26	0.96	0.96
13	ROB	0.44	0.94	0.94	0.38	0.95	0.95	0.34	0.95	0.95
					C_1					
5	CL	3.19	0.21	0.21	3.19	0.21	0.21	3.19	0.21	0.21
5	ROB	0.79	0.82	0.82	0.79	0.82	0.82	0.79	0.82	0.82
9	CL	3.22	0.21	0.21	3.21	0.21	0.21	3.21	0.21	0.21
9	ROB	1.06	0.79	0.79	1.00	0.80	0.80	0.96	0.80	0.80
13	CL	3.22	0.21	0.21	3.21	0.21	0.21	3.21	0.21	0.21
13	ROB	1.09	0.79	0.79	1.01	0.79	0.80	0.97	0.80	0.80
					C_2					
5	CL	3.02	0.26	0.26	3.02	0.26	0.26	3.02	0.26	0.26
5	ROB	1.02	0.78	0.77	1.04	0.77	0.77	1.02	0.78	0.78
9	CL	3.08	0.26	0.26	3.06	0.26	0.26	3.05	0.26	0.26
9	ROB	1.41	0.72	0.72	1.30	0.74	0.74	1.26	0.74	0.74
13	CL	3.08	0.26	0.26	3.07	0.26	0.26	3.06	0.26	0.26
13	ROB	1.43	0.72	0.72	1.34	0.73	0.73	1.28	0.74	0.74

Table S.1: MISE and mean over replications of $\widehat{\rho}_X = \widehat{\rho}_{\text{CL},XX,\text{CLEAN}}(\phi_1, \widetilde{\phi}_1)$ and $\widehat{\rho}_Y = \widehat{\rho}_{\text{CL},YY,\text{CLEAN}}(\psi_1, \widetilde{\psi}_1)$ for SCCA in the basis expansion domain (when using the Fourier basis) and different contamination settings. The dimension of the subspaces considered are $d = 5, 9, 13$ and the selected values for the smoother parameter are $\tau = 20, 30, 40$.

Penalty parameter		$\tau = 20$			$\tau = 30$			$\tau = 40$		
d	Estimator	MISE	$\widehat{\rho}_X$	$\widehat{\rho}_Y$	MISE	$\widehat{\rho}_X$	$\widehat{\rho}_Y$	MISE	$\widehat{\rho}_X$	$\widehat{\rho}_Y$
					C_0					
5	CL	0.13	0.97	0.97	0.13	0.97	0.97	0.13	0.97	0.97
5	ROB	0.18	0.96	0.96	0.18	0.96	0.96	0.18	0.96	0.96
9	CL	0.30	0.96	0.96	0.26	0.96	0.96	0.24	0.96	0.96
9	ROB	0.38	0.95	0.95	0.34	0.95	0.95	0.31	0.95	0.95
13	CL	0.33	0.96	0.96	0.28	0.96	0.96	0.26	0.96	0.96
13	ROB	0.43	0.94	0.94	0.37	0.95	0.95	0.33	0.95	0.95
20	CL	0.34	0.96	0.96	0.29	0.96	0.96	0.26	0.96	0.96
20	ROB	0.44	0.94	0.94	0.38	0.95	0.94	0.34	0.95	0.95
					C_1					
5	CL	3.19	0.21	0.21	3.19	0.21	0.21	3.19	0.21	0.21
5	ROB	0.79	0.82	0.82	0.78	0.82	0.82	0.78	0.82	0.82
9	CL	3.21	0.21	0.21	3.21	0.21	0.21	3.21	0.21	0.21
9	ROB	1.03	0.79	0.79	0.98	0.80	0.80	0.96	0.80	0.80
13	CL	3.22	0.21	0.21	3.21	0.21	0.21	3.21	0.21	0.21
13	ROB	1.08	0.79	0.79	1.02	0.79	0.79	0.98	0.80	0.80
20	CL	3.22	0.22	0.21	3.22	0.21	0.21	3.20	0.21	0.21
20	ROB	1.08	0.79	0.79	1.03	0.79	0.79	0.98	0.80	0.80
					C_2					
5	CL	3.01	0.26	0.26	3.01	0.26	0.26	3.01	0.26	0.27
5	ROB	0.99	0.78	0.78	0.98	0.78	0.78	0.97	0.79	0.79
9	CL	3.06	0.26	0.26	3.05	0.26	0.26	3.04	0.27	0.27
9	ROB	1.29	0.74	0.75	1.22	0.75	0.76	1.19	0.76	0.76
13	CL	3.07	0.27	0.27	3.05	0.27	0.27	3.04	0.27	0.27
13	ROB	1.32	0.74	0.74	1.24	0.75	0.76	1.17	0.77	0.76
20	CL	3.07	0.27	0.27	3.07	0.26	0.26	3.04	0.27	0.27
20	ROB	1.35	0.74	0.74	1.28	0.74	0.75	1.24	0.75	0.75

Table S.2: MISE and mean over replications of $\widehat{\rho}_X = \widehat{\rho}_{\text{CL},XX,\text{CLEAN}}(\phi_1, \widetilde{\phi}_1)$ and $\widehat{\rho}_Y = \widehat{\rho}_{\text{CL},YY,\text{CLEAN}}(\psi_1, \widetilde{\psi}_1)$ for SCCA in the basis expansion domain (when using the cubic B-splines) and different contamination settings. The dimension of the subspaces considered are $d = 5, 9, 13, 20$ and the selected values for the smoother parameter are $\tau = 20, 30, 40$.

Penalty parameter	$\tau = 20$			$\tau = 30$			$\tau = 40$		
Estimator	MISE	$\widehat{\rho}_X$	$\widehat{\rho}_Y$	MISE	$\widehat{\rho}_X$	$\widehat{\rho}_Y$	MISE	$\widehat{\rho}_X$	$\widehat{\rho}_Y$
C_0									
CL	0.40	0.95	0.95	0.38	0.95	0.95	0.38	0.94	0.94
ROB	0.53	0.92	0.92	0.51	0.92	0.92	0.51	0.92	0.92
C_1									
CL	3.16	0.23	0.23	3.14	0.23	0.23	3.12	0.24	0.24
ROB	1.16	0.77	0.77	1.11	0.77	0.77	1.09	0.77	0.77
C_2									
CL	3.17	0.24	0.24	3.17	0.23	0.23	3.17	0.23	0.23
ROB	1.43	0.72	0.72	1.35	0.73	0.73	1.32	0.73	0.73

Table S.3: MISE and mean over replications of $\widehat{\rho}_X = \widehat{\rho}_{\mathrm{CL},XX,\mathrm{CLEAN}}(\phi_1, \widetilde{\phi}_1)$ and $\widehat{\rho}_Y = \widehat{\rho}_{\mathrm{CL},YY,\mathrm{CLEAN}}(\psi_1, \widetilde{\psi}_1)$ for SCCA and different contamination settings. The selected values for the smoother parameter are $\tau = 20, 30, 40$.

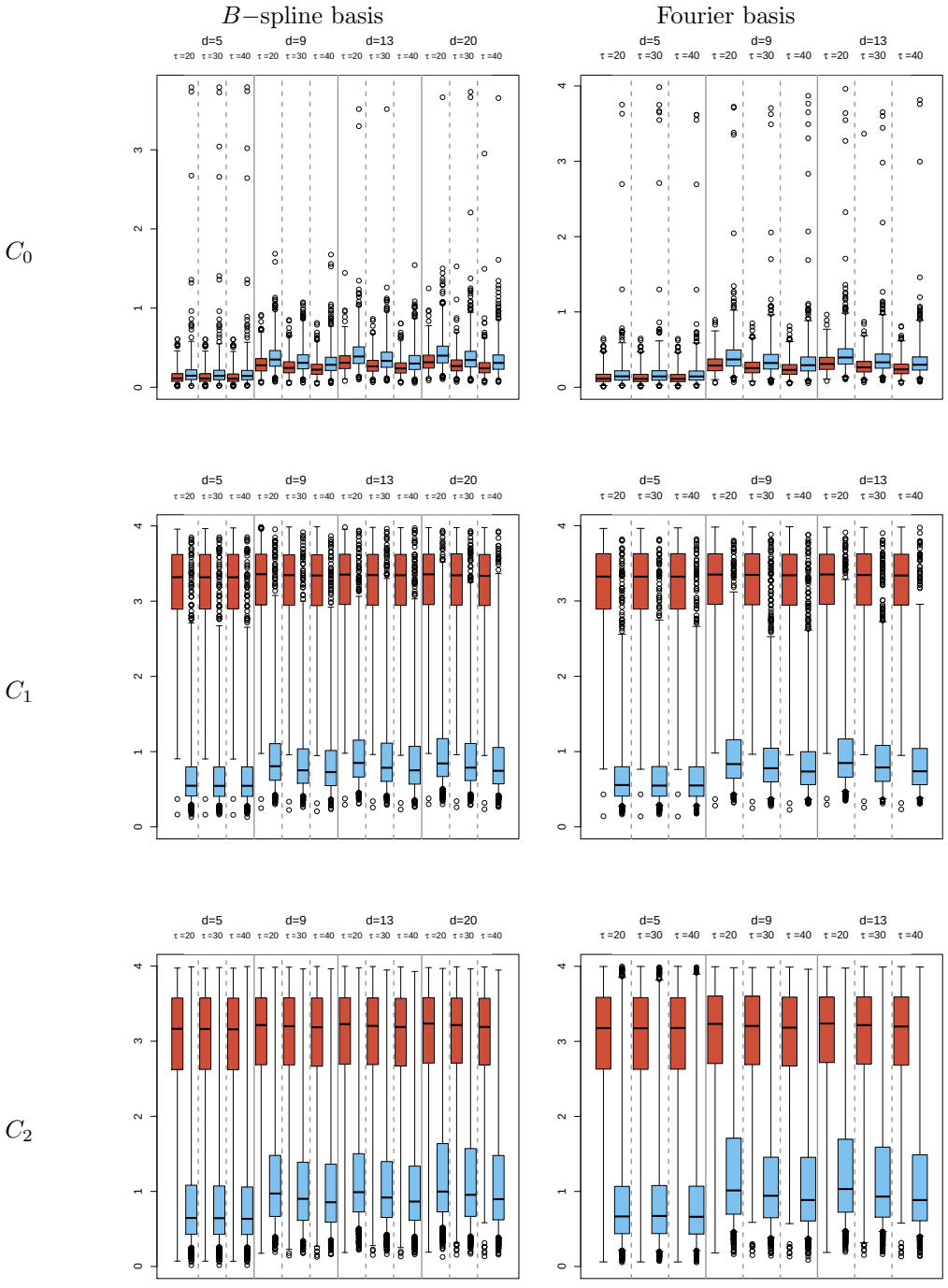

Figure S.2: Adjusted boxplots of MISE. The red boxes correspond to the classical procedure, while the light blue ones to the robust one.

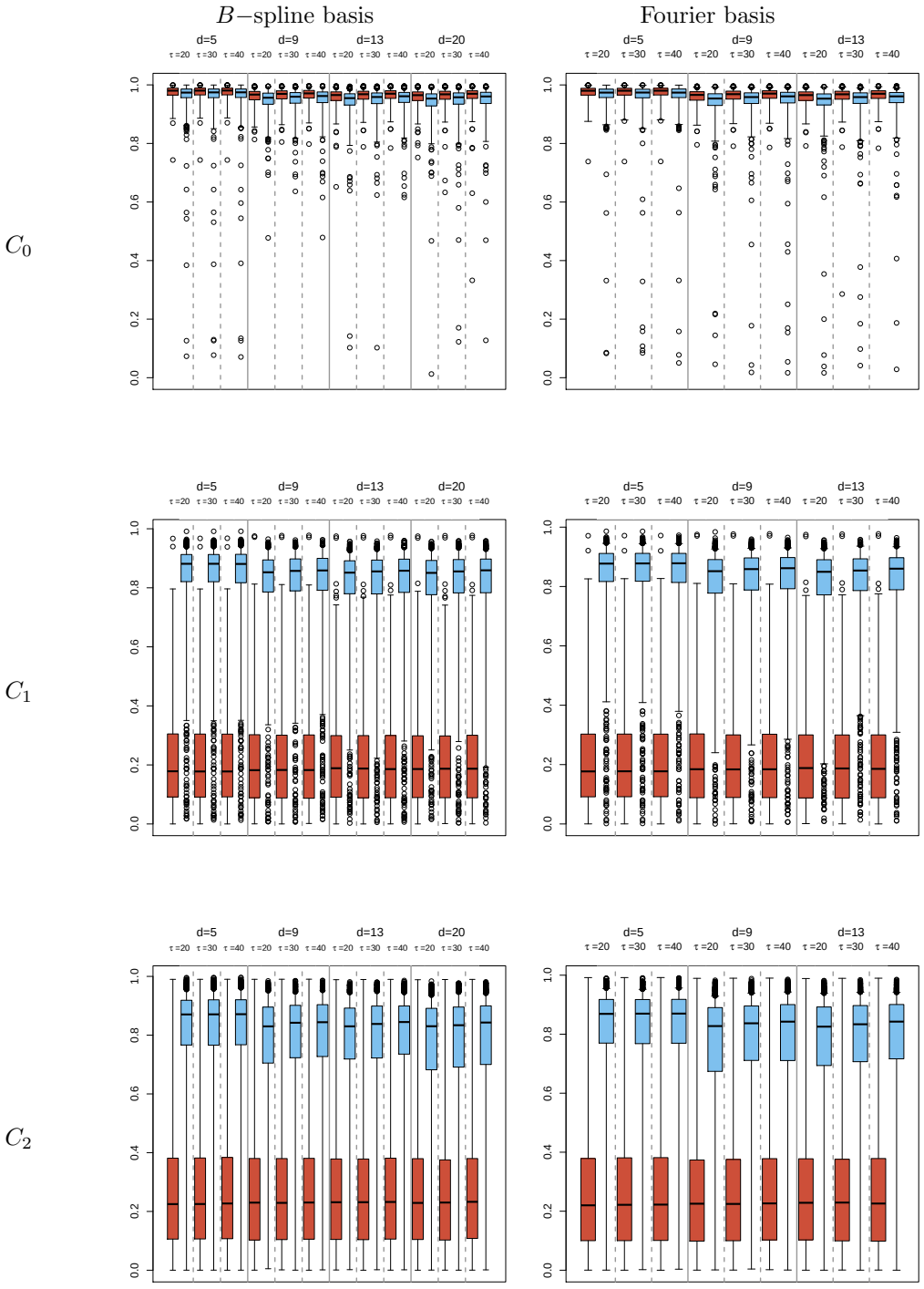

Figure S.3: Adjusted boxplots of $\widehat{\rho}_X$. The red boxes correspond to the classical procedure, while the light blue ones to the robust one.

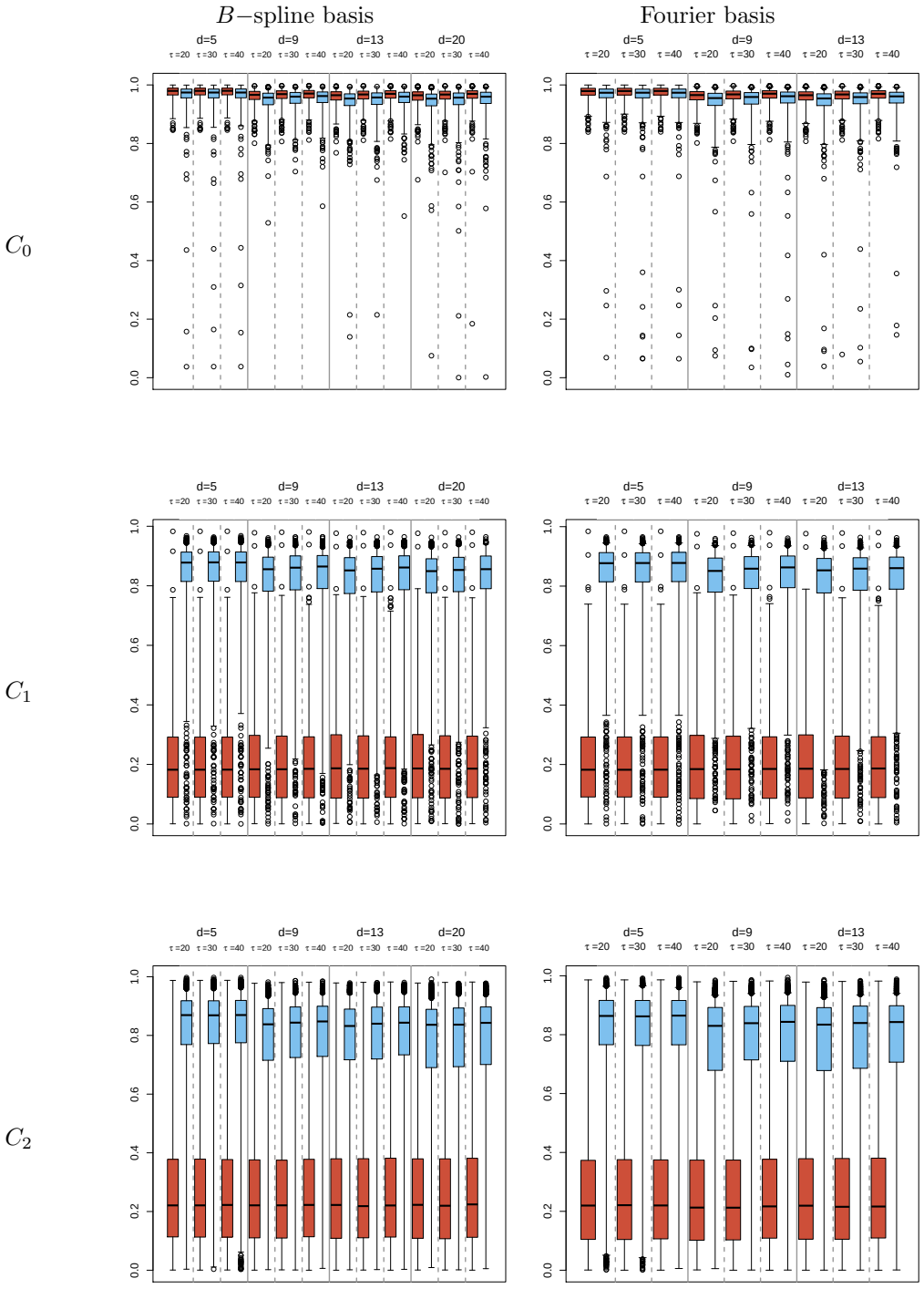

Figure S.4: Adjusted boxplots of $\widehat{\rho}_Y$. The red boxes correspond to the classical procedure, while the light blue ones to the robust one.